

Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from India

ReadMe

These semi-structured interview data are part of the project “Research Integrity and Privacy” of the HEADT Centre (<https://headt.eu>), a study on the presumed conflict between publishing of research data and the issues of data protection.

Researcher:

Katarzyna Biernacka
Humboldt-Universität zu Berlin
Department of Computer Science
Computer Science Education | Computer Science and Society
biernack@hu-berlin.de
ORCID: 0000-0002-6363-0064

Project members:

Principal Investigator: Prof. Niels Pinkwart
Student assistants: Malte Smits and Rudi Rudiak

Files:

- data_dictionary.csv
Overview of the datasets
- os_025.docx and os_025.pdf
Semi-structured interview with a researcher from climate impact research in India
- os_026.docx and os_026.pdf
Semi-structured interview with a researcher from climate impact research in India
- os_027.docx and os_027.pdf
Semi-structured interview with a researcher from medicine in India
- os_028.docx and os_028.pdf
Semi-structured interview with a researcher from climate impact research in India
- os_032.docx and os_032.pdf
Semi-structured interview with a researcher from learning analytics in India
- questions.pdf
Questions used for the semi-structured interviews
- transcription_norms.pdf
Norms and rules that have been applied for transcription



1. Data collection

The semi-structured interviews included in this data set were collected between November 2019 and January 2020 with scientists from India. Researchers from learning analytics, medicine and climate impact research were interviewed in order to assess their behaviour regarding the publishing of their research data (see **questions.pdf** for more details). Both junior (2-4 years of experience) and senior scientists (more than 5 years of experience) were included in the semi-structured interviews.

The semi-structured interviews were conducted 1) in person and recorded with a TASCAM DR-40 recorder, or 2) via DFNconf video conference system with the camera switched off. In the latter case the audio was extracted to an m4a-file in the next step. The audio files, in both cases, are preserved at the TSM service of the Humboldt-Universität zu Berlin, where data access is protected and regulated via a login. Third parties have no access to the data.

2. Data processing

The produced audio files were transcribed by two student assistants (see project members) using the denaturalized technique in order to preserve the purest and least altered form of the interview. In this way the transcription is as transparent as possible. The transcription norms and rules defined in the **transcription_norms.pdf** were adopted.

The anonymisation was carried out manually by the researcher and subsequently checked by the project members. In most cases, it was decided to aggregate the information, e.g. “university” or “India” (instead of the exact institute name or city name). Where it was not possible otherwise (because the risk of de-identification was too high) strong abstraction was used (e.g. “name”).

3. Data evaluation

The qualitative evaluation of the semi-structured interviews is based on the Grounded Theory methodology. These hypotheses shall be verified and quantified via a wide spread online survey that will follow.

